Article

Not peer-reviewed version

Sound of Surveillance: Enhancing Machine Learning-Driven Drone Detection with Advanced Acoustic Augmentation

Sebastian Kümmritz *

Posted Date: 30 January 2024

doi: 10.20944/preprints202401.2114.v1

Keywords: UAV classification; machine learning; audio data augmentation; UAV detection

Article

Sound of Surveillance: Enhancing Machine Learning-Driven Drone Detection with Advanced Acoustic Augmentation

Sebastian Kümmritz

H2 Think gGmbH, kuemmritz@h2think.org

Abstract: In response to the growing challenges in drone security and airspace management, this study introduces an advanced drone classifier, capable of detecting and categorizing Unmanned Aerial Vehicles (UAVs) based on acoustic signatures. Utilizing a comprehensive database of drone sounds across EU-defined classes (C0 to C3), this research leverages machine learning (ML) techniques for effective UAV identification. The study primarily focuses on the impact of data augmentation methods—pitch shifting, time delays, harmonic distortion, and ambient noise integration—on classifier performance. These techniques aim to mimic real-world acoustic variations, thus enhancing the classifier's robustness and practical applicability. Results indicate that moderate levels of augmentation significantly improve classification accuracy. However, excessive application of these methods can negatively affect performance. The study concludes that sophisticated acoustic data augmentation can substantially enhance ML-driven drone detection, providing a versatile and efficient tool for managing drone-related security risks. This research contributes to UAV detection technology, presenting a model that not only identifies but also categorizes drones, underscoring its potential for diverse operational environments.

Keywords: UAV detection; UAV classification; machine learning; audio data augmentation

1. Introduction

The rapid proliferation of UAVs, commonly known as drones, has opened a spectrum of opportunities and applications ranging from aerial photography to logistics. However, this growth has also ushered in significant challenges in airspace management and security, highlighted by incidents like the disruptions at London Gatwick Airport in 2018 [1]. These challenges necessitate the development of effective drone detection and classification systems.

While traditional methods like radar [2], RF-based techniques [3], and visual systems [4] are prevalent, their drawbacks in terms of cost, range, and environmental sensitivity are well-recognized. Recent years have seen a growing interest in acoustic-based drone detection, a field that has evolved rapidly due to its cost-effectiveness and operational flexibility. Numerous studies have explored the use of acoustic signatures for UAV detection, underscoring the potential of this approach [5–8].

However, it is crucial to acknowledge that each detection technique, including acoustic-based methods, has its inherent advantages and disadvantages. No single technique suffices in creating a comprehensive and effective drone detection system. As Park et al. aptly noted, relying solely on one method of detection inevitably leads to gaps in drone detection capabilities, posing challenges in successfully neutralizing illegal drones [9]. This paper focuses primarily on acoustic detection due to its cost efficiency. The use of small, cost-effective detection devices equipped with MEMS microphones could be widely deployed in sensor networks, potentially compensating for some of the limitations inherent in acoustic-based detection. By integrating these devices into extensive networks, a more thorough and efficient detection framework can be established, leveraging the scalability and economic feasibility of acoustic technology.

Building on prior work, 'Comprehensive Database of Drone Sounds for Machine Learning' [10], a substantial open-access database of drone audio data has been developed. This database, meticulously

 doi:10.20944/preprints202401.2114.v1

compiled and categorized, covers a range of UAV classes from C0 to C3. An extensive collection of 44 different drone models is included, encompassing a significant total duration of 23.42 hours of recordings. This comprehensive assembly of data forms a robust foundation for the development and training of ML algorithms for drone detection.

The focus of this paper is to introduce sophisticated drone classifiers capable of not only detecting drones but also classifying them into distinct categories (C0 to C3) as defined by the European Union's drone regulations [11]. These regulations classify UAVs based on their weight, capabilities, and intended use. Please refer to Table 1 for a detailed breakdown of these categories, which provides a concise overview of the classification and characteristics of each category as per the EU drone regulations. This breakdown is essential for understanding the diverse range of UAVs that the classifiers can detect and categorize.

Table 1. Overview of EU Drone Categories

Category	Description
C0	Drones weighing less than 250 grams, typically for leisure and recreational use.
C1	Small drones weighing less than 900 grams, used for both recreational and commercial purposes, with more features than C0 drones.
C2	Drones weighing less than 4 kilograms, used for complex commercial operations, requiring advanced operational skills.
C3	Larger drones weighing less than 25 kilograms, generally used for specialized commercial tasks demanding specific capabilities.

Our research primarily investigates the impact of data augmentation on the performance of these classifiers. The classifiers were initially trained using pristine drone sound recordings obtained in an anechoic chamber, ensuring the purity and clarity of the audio data. Subsequently, we explored various augmentation techniques to simulate real-world environmental conditions, thereby enhancing the classifiers robustness and applicability in diverse scenarios.

Data augmentation emerges as a powerful tool, especially when dealing with limited training data. As noted in [12], one of the main drawbacks of deep learning approaches is the need for a large amount of training data. According to [13] one solution could be the augmentation of existing data. However, there is not a single data augmentation protocol that outperforms all the others in all the tests [13]. It is important to recognize that specific augmentation approaches must be carefully tailored to the dataset in question. For instance, [14] showed that augmenting their spectrograms by translations, adding random noise, and multiplying the input by a random value close to one, did not significantly improve their classification of marmoset audio signals.

The potential of audio-based ML systems in UAV detection is underscored by this approach, presenting a cost-effective, scalable, and versatile solution to the myriad challenges posed by the widespread use of drones. The specifics of the neural network architecture, training methods, and augmentation techniques utilized in the study are explored in the methodology section. Additionally, an overview of two major data collection campaigns, forming the bulk of the training and validation data, is provided in this section.

The results section provides a comprehensive analysis of various augmentation techniques, examining their effects both independently and collectively on the performance of multiple classifiers. Additionally, this section aims to explore the real-world applicability of these classifiers through an experimental deployment. The data from an outdoor experiment, conducted in a diverse acoustic environment, are intended to demonstrate the potential adaptability and effectiveness of the classifiers developed in this study.

By employing this methodical approach, the paper seeks to detail the process of refining and evaluating different classifiers, each characterized by its distinct features and performance metrics. This ongoing process of development and meticulous testing is designed to thoroughly assess each

 doi:10.20944/preprints202401.2114.v1

classifier's efficiency and reliability in practical scenarios. Such a careful and focused examination of classifiers in an actual outdoor setting aims to significantly enhance our understanding and capabilities in UAV detection and classification, providing valuable insights into the potential practical applications of these systems.

2. Materials and Methods

An essential aspect of the research methodology is a commitment to transparency and reproducibility. In pursuit of this, the source code used to achieve the study's results has been made publicly available on GitHub [15]. This code, along with the audio data from the database (refer to subsection 2.1), equips fellow researchers and interested parties with the necessary tools to replicate the findings.

This approach ensures that the methods and results can be independently verified, thus fostering a collaborative and open scientific environment. Providing access to both the source code and the audio data is intended to facilitate a comprehensive understanding and potential further development of the research within the academic community.

2.1. Data

The training data for the drone classifier predominantly originates from two extensive measurement campaigns, as detailed in the prior work [10]. The initial campaign involved recordings in an anechoic chamber, providing high-quality, reflection-free audio samples from various drone models. These recordings were crucial in establishing a clean acoustic profile baseline for the initial training stages of the ML model.

The subsequent phase of the study involved an outdoor experiment at the Fraunhofer IVI test oval in Dresden. This experiment is illustrated in Figure 1 using an OpenStreetMap graphic, which specifically showcases the trajectory of a drone during one of the measurement sessions. In this graphic, green dots represent the locations of microphones, strategically positioned to capture the acoustic data from the drone. The trajectory of the drone, predominantly depicted in red, reflects its constant altitude during flight and is presented as a visually continuous line rather than discrete points. This continuous appearance is due to the short intervals at which the drone's GPS positions were logged.

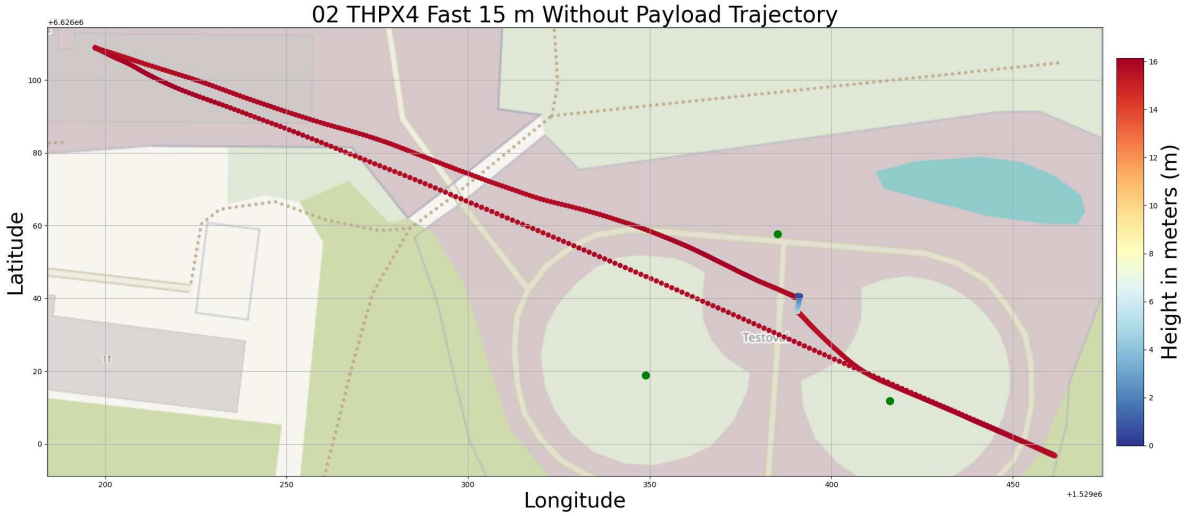


Figure 1. Visualization of a drone's flight path over the Fraunhofer IVI test oval, with color-coded altitude indicators and microphone positions.

The trajectory was intentionally designed to challenge the classifier's detection capabilities by having the drone initially move to a minimally audible distance and then return. This tested the

 doi:10.20944/preprints202401.2114.v1

classifier's range and its proficiency in discerning drone sounds amidst varying real-world background noises, thus enhancing the system's robustness and practical deployment readiness.

For targeted data preparation for training and validation, we employed a SQL query to categorize and retrieve the data from our data base (the exact query can be found in [15]). The 'Training' folder comprised all drone classes from anechoic chamber measurements, organized into subfolders C0, C1, C2, and C3, corresponding to different drone classifications. A 'Validation' folder contained the remaining drone data. Additionally, a 'no drone' folder was created with audio files from the database where drones were inaudible. The limited 'no drone' samples necessitated supplementing the dataset with external sources like YouTube and https://www.salamisound.de/, incorporating diverse environmental sounds to refine the model's distinguishing capabilities.

For a comprehensive understanding of the measurement methods and outcomes, we direct readers to [10]. The complete dataset is principally available at https://mobilithek.info/offers/605778370199691264, but the platform does have certain data management constraints. The data must be downloaded and converted from the platform, with guidance provided in the linked material. After the conversion process, the data can be locally hosted as an SQL database. Alternatively, readers can request a download link from the author where the data has been readily formatted for an SQL database.

2.2. Network and Training

In the development of the audio-based drone detection model, the vggish network, a neural network architecture specifically designed for acoustic applications, was utilized [16]. Known for its effectiveness in processing and analyzing complex sound data, the vggish network presents a good solution for the classification of drone sounds.

Structured to accept Mel-Frequency Cepstral Coefficients (MFCCs) matrices, the input layer of the vggish network in this model accommodates formats of 96 x 64 x n, with 'n' representing a variable number of consecutive MFCCs. This particular format is selected for its capacity to encapsulate the essential characteristics of sound in a compact representation, greatly aiding in the processing and recognition of distinct acoustic signatures of various drone models.

As cited in [17], "MFCCs offer a compact description of the spectral envelope of sound in the cepstral domain." Employed extensively in audio signal processing, especially for voice and sound recognition, MFCCs efficiently encapsulate the short-term power spectrum of a sound. Within the scope of this project, the MFCC matrices furnish a substantial yet manageable dataset, which the vggish network processes to achieve precise identification and classification of drone sounds in diverse acoustic settings.

In the study conducted, two types of classifiers were developed to improve the efficiency and accuracy of the drone detection system. The first classifier was used to distinguish between 'drone' and 'no drone' sounds, while the second classifier was designed to categorize identified drone sounds into one of the four drone categories (C0, C1, C2, C3). This approach with two classifier types addresses two main aspects:

1. **Robustness in Drone Category Classification:** The distinction between sounds from different drone categories is more subtle compared to the distinction between drone and non-drone sounds. By first excluding 'no drone' sounds, the training of the drone category classifier becomes more robust, concentrating solely on the nuanced differences between the drone classes.
2. **Efficiency in Operational Deployment:** Considering that in a real-world deployment scenario, drone sounds are expected to be less frequent than non-drone sounds, a cascaded approach is more practical. Initially, the system continuously monitors for the presence of drones. Once a drone is detected, the second classifier steps in to determine its category. This method is particularly beneficial when integrating neuromorphic technology like SynSense's Xylo [18], which offers binary classifications with ultra-low power consumption. In a drone detection system, detecting

 doi:10.20944/preprints202401.2114.v1

a drone could trigger a wake-up process for a more power-intensive unit (e.g., an Arduino) to perform the detailed classification and relay the detection to the cloud via a 5G connection.

To optimize the model, consistent training parameters were maintained throughout all investigations. Stochastic Gradient Descent with Momentum (sgdm) was chosen as the optimization algorithm, playing a crucial role in effectively balancing the convergence rate and the accuracy of the model. The learning rate was managed using a piecewise schedule, starting at an initial rate of 0.001. This approach was integral in facilitating efficient learning over epochs while preventing overfitting. The learning rate was designed to drop by a factor of 0.1 every three epochs, ensuring that the model gradually refined its focus on learning without missing finer details.

In terms of data processing, training was conducted in batches, each containing 256 samples. The chosen batch size balanced computational efficiency with the model's capacity to learn from a diverse range of data samples within each epoch. Furthermore, the training was limited to a maximum of 12 epochs. This specific number of epochs was established as the most effective for the model's training on the dataset, facilitating comprehensive learning while avoiding unnecessary computational strain.

2.3. Augmentation Techniques and Data Preperation

In the process of enhancing the robustness of the drone sound classification model, a variety of audio augmentation techniques were explored. The primary methods focused on included pitch shifting, adding delay, introducing harmonic distortions, and incorporating background noise. These techniques were selected for their ability to simulate real-world acoustic variations, thus equipping the model to effectively operate in diverse environmental conditions.

1. Pitch Shifting: The pitch of drone sounds was altered without changing the playback speed, simulating variations in drone motor speeds. In the Matlab code, the applyCustomPitching function (see Appendix 1) shifts the pitch randomly within a specific range, broadening the classifier's ability to recognize different drone sounds.
2. Adding Delay: The applyDelay function (see Appendix 2) introduces a time delay to the original sound. This delay, varied in length and amplitude, mimics echo effects in various environments, enhancing the model's adaptability to different acoustic settings.
3. Environmental Noise: The applyEnvironmentalNoise function (see Appendix 3) was used to mix ambient noises into the drone sounds. Ambient sounds, sourced from diverse environments, were employed to train the model in differentiating drone noises from background sounds in real-world scenarios.
4. Harmonic Distortion: The applyHarmonicDistortion function (see Appendix 4) was utilized to add harmonic distortions, simulating the effect of sound traveling through different media. This technique challenges the model to maintain accuracy in complex acoustic landscapes.

It should be noted that the augmentation techniques were exclusively applied to the training data and not to the validation data. This approach is based on the principle, as suggested by [19], that augmented data might not accurately represent realistic scenarios or could introduce alteration artifacts. As stated, "Depending on the domain and the specific alterations applied during augmentation, the augmented data does not necessarily perfectly resemble realistic data or may show alteration artifacts. Therefore, data augmentation is usually only applied to the training data and not to the validation or test data." This method ensures that while the model is trained on a varied and challenging dataset, its performance is evaluated using unaltered, real-world data, thus providing a realistic assessment of its capabilities.

In the data preparation phase, the audio data was initially segmented into one-second chunks. This decision was influenced by the findings presented in [6], which suggested through heuristic observations that one-second audio clips are optimal for drone detection.

Additionally, a pre-classification step was implemented to remove any segments of the audio signal that did not contain drone sounds. This was achieved by evaluating the Harmonic-to-Noise

 doi:10.20944/preprints202401.2114.v1

Ratio in the signal. Since the recordings in the anechoic chamber contained exclusively drone noises, with minimal other sounds present except for microphone noise or occasional sounds made by experimenters, this method proved effective in isolating relevant drone sounds.

Furthermore, each audio chunks was normalized to 90 percent amplitude. This normalization process aims to minimize the impact of volume variations, such as those caused by the distance of the drone from the microphone, on the classification outcome. This step ensures a more consistent input level across all data used for training the classifier.

A further step in the data preparation process involved applying a bandpass filter to all data, limiting the frequency range to between 200 Hz and 20,000 Hz. This approach was adopted based on the observation that significant signal components were absent below 200 Hz and above 20,000 Hz. The application of this filter, by excluding irrelevant frequencies, not only improved the overall signal quality but also enhanced the ability of the classification model to accurately identify and categorize drone sounds.

3. Results

3.1. 'Drone' vs. 'no Drone' classification

The dataset was initially split evenly into training and validation sets to assess the classifier's ability to distinguish between drone and non-drone sounds. The training data for drone sounds underwent augmentation with techniques such as pitching (maxPitch = 0.085), delay (maxDelay = 21 ms; maxAmplitude = 30%), and harmonic distortions (distortionLevel = 11.1%). These augmentation parameters were selected based on an initial iteration of our augmentation studies, aiming to balance between enhancing the classifier's adaptability and maintaining the integrity of the drone sounds. Further details on these augmentation techniques will be provided later.

The confusion matrix for the 'Drone' vs. 'no Drone' classifier suggests a drone detection accuracy of 99.1% and a non-drone detection accuracy of 97.2%. These results indicate a high degree of reliability in differentiating between drone and non-drone acoustic signatures. However, it is important to note that the augmentation parameters used here represent a preliminary selection. Further fine-tuning of these parameters could potentially improve the classifier's performance. For instance, adjustments in the level of pitch, delay, or distortion might refine the classifier's sensitivity to subtle variations in drone sounds, particularly in challenging acoustic environments.

Despite the potential for further optimization, the current iteration of the 'Drone' vs. 'no Drone' classifier has demonstrated robust performance. Subsequent investigations, as discussed in later sections, have reinforced the reliability of this classifier in various scenarios. This consistency underscores the effectiveness of the chosen augmentation techniques and parameters, even in their initial form. It suggests that the classifier, as developed, can serve as a solid foundation for future enhancements and refinements.

3.2. Drone class classification without augmentation

3.2.1. Classifier Variability and Stochastic Processes

In an effort to establish a baseline for drone sound classification, four distinct classifiers were trained without the application of any data augmentation techniques, under ostensibly identical conditions. Notably, even with the same script executed in each run, significant variations in the outcomes were observed across the classifiers. The confusion matrices depicted in Figure 2 provide a clear visual representation of these discrepancies. For instance, the accuracy for classifying the C0 category varied from 91.4% to 96.6%, while the C1 category showed a fluctuation from 92.7% to 96.7%. Such variations were also present in the more nuanced classifications of the C2 and C3 categories.

 doi:10.20944/preprints202401.2114.v1

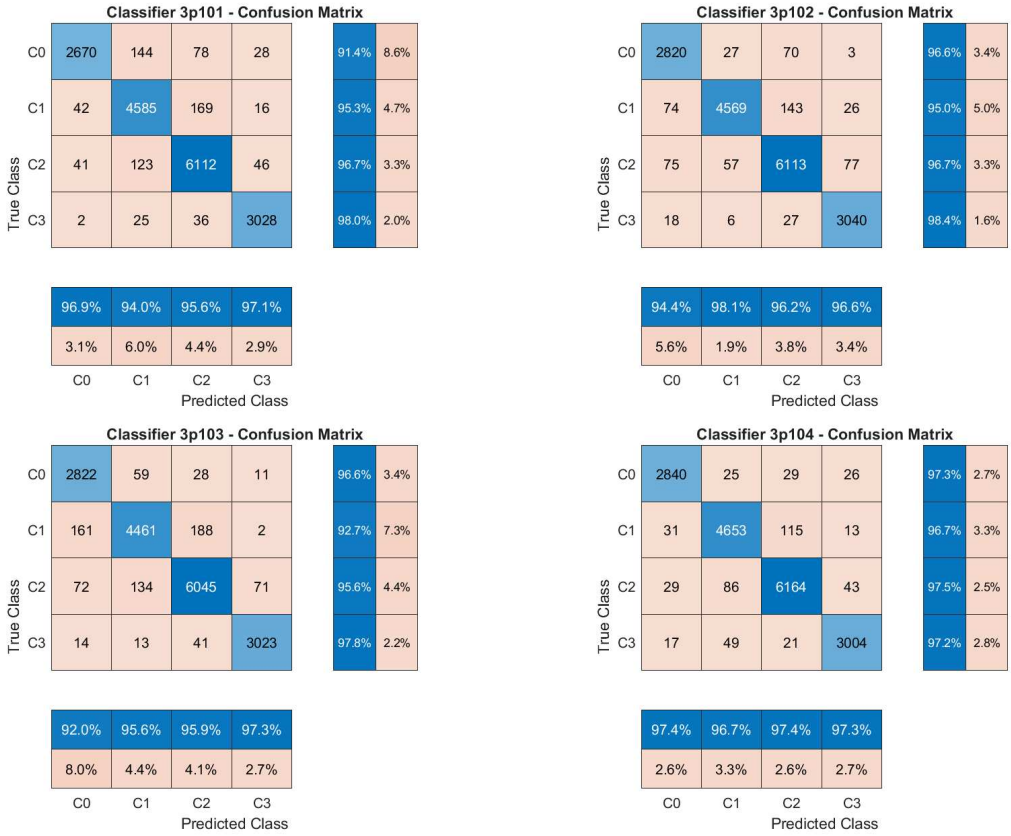


Figure 2. Confusion matrices for four different classifiers (without augmentation), trained under identical conditions.

These inconsistencies can largely be attributed to the stochastic nature of processes involved in ML model training, such as the random initialization of weights and the probabilistic elements inherent in the learning algorithms. The differences in the results underscore the significant influence that these random processes can have on the performance and generalization ability of ML models. The findings from this baseline experiment without augmentation emphasize the importance of accounting for these random processes when training models, as they can lead to considerable variability in performance despite identical training setups. They also highlight the need for rigorous experimental design, such as setting fixed seeds for random number generators to ensure reproducibility and reliability in ML research.

Table 2 presents a summary of the performance of multiple classifiers trained under identical conditions without data augmentation. The classifiers are divided into four groups based on the random seed initialization: no seed, seed initialized to 1, seed initialized to 2, and seed initialized to 3.

 doi:10.20944/preprints202401.2114.v1

Table 2. Classifier Performance Comparison without Augmentation: Accuracy Metrics and Resulting Variance with and without Seed Initialization

Seed		True Class					Predicted Class				
		C0	C1	C2	C3	mean	C0	C1	C2	C3	mean
	no	91.4%	95.3%	96.7%	98.0%		96.9%	94.0%	95.6%	97.1%	
	no	96,6%	95,0%	96,7%	98,4%		94,4%	98,1%	96,2%	96,6%	
	no	96,6%	92,7%	95,6%	97,8%		92,0%	95,6%	95,9%	97,3%	
	no	97,3%	96,7%	97,5%	97,2%		97,4%	96,7%	97,4%	97,3%	
std		2,4%	1,4%	0,7%	0,4%	**1,2%**	2,2%	1,5%	0,7%	0,3%	**1,2%**
	1	96,5%	98,2%	97,7%	99,2%		98,1%	98,0%	97,5%	98,5%	
	1	98,1%	98,5%	97,0%	98,7%		95,3%	98,3%	98,3%	99,0%	
	1	92,7%	98,2%	97,4%	99,1%		96,3%	97,0%	97,3%	97,6%	
	1	97,1%	97,6%	96,5%	99,2%		95,7%	98,2%	97,5%	97,5%	
std		2,0%	0,3%	0,5%	0,2%	**0,8%**	1,1%	0,5%	0,4%	0,6%	**0,6%**
	2	96,8%	96,0%	97,7%	98,1%		94,6%	96,5%	97,9%	99,0%	
	2	95,5%	98,2%	97,9%	99,0%		98,5%	97,2%	97,6%	98,3%	
	2	97,0%	96,6%	97,9%	99,2%		96,0%	97,7%	97,8%	98,6%	
	2	93,3%	98,3%	97,0%	98,0%		98,7%	94,4%	97,6%	98,0%	
std		1,5%	1,0%	0,4%	0,5%	**0,8%**	1,7%	1,3%	0,1%	0,4%	**0,9%**
	3	95,5%	96,7%	96,2%	99,1%		97,3%	96,6%	96,5%	97,0%	
	3	94,8%	97,6%	97,3%	98,7%		97,5%	97,1%	97,1%	97,2%	
	3	96,2%	95,7%	97,3%	99,3%		97,4%	97,8%	96,0%	97,7%	
	3	96,4%	96,4%	96,5%	98,4%		96,4%	96,2%	96,5%	97,8%	
std		0,6%	0,7%	0,5%	0,3%	**0,5%**	0,4%	0,6%	0,4%	0,3%	**0,4%**

For the group without seed initialization, the mean accuracy across classes (C0 to C3) exhibits notable fluctuations, evidenced by a higher standard deviation. This variation in classifier performance underscores the impact of uncontrolled random processes. Specifically, the 'no seed' group demonstrates more considerable variability in outcomes, with an average standard deviation of 1.2% for both true and predicted class categories, suggesting less consistent classifier performance. In contrast, the groups with seed initialization show a markedly reduced standard deviation in accuracy for both true and predicted class categories. For these seeded groups, the average standard deviation is lower, at around 0.9% in maximum and at around 0.3% in minimum, indicating a more even and consistent performance across multiple runs of the classifiers.

3.2.2. Real-World Performance of Non-Augmented Classifiers

The performance of the first drone classifier, with seed 1 as detailed in Table 2, was evaluated in classifying a C3 drone from an outdoor experiment. The outcomes of this real-world application are depicted in Figure 3.

The spectrogram in the upper section of Figure 3 clearly illustrates a typical acoustic footprint of the drone's activity. The typical acoustic spectrum of a drone is characterized by a distinctive pattern of harmonics across mid to high frequencies, often with peaks in lower frequencies generated by the rotors and motors. The drone initiates movement at around 7.5 seconds, with a stationary phase until approximately 25.5 seconds, and subsequently moves away from the microphone. Its farthest distance from the microphone, where the acoustic signature is weakest, is reached around 52 seconds before it begins its return journey. The drone passes directly overhead at 66.5 seconds and finally lands at 109 seconds.

 doi:10.20944/preprints202401.2114.v1

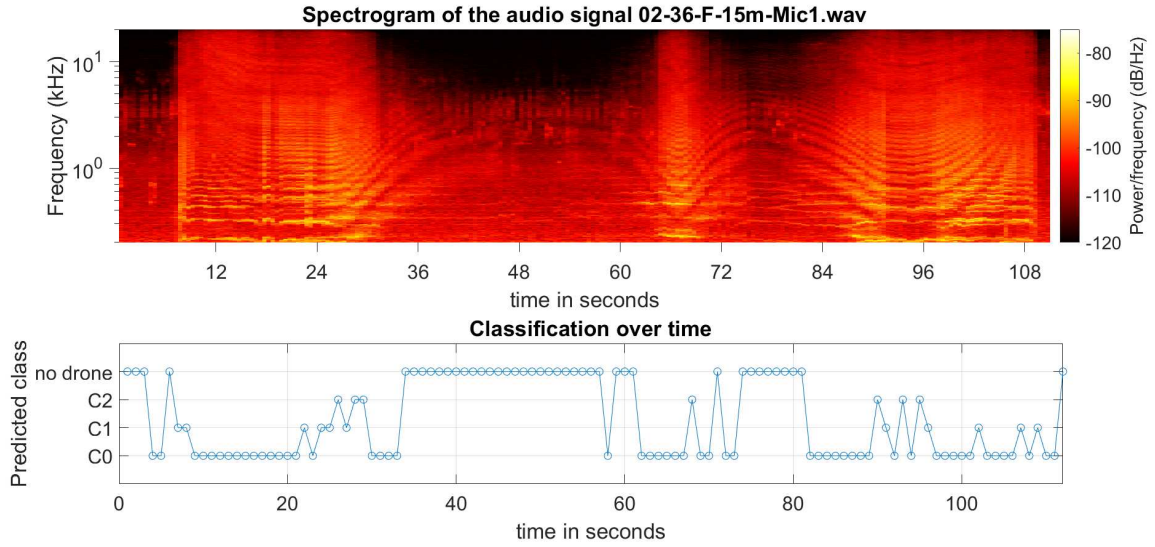


Figure 3. Top: Spectrogram of the audio signal capturing the drone's acoustic signature during the outdoor experiment. Bottom: Classification results over time, showing the classifier's predictions.

The classifier's predictions over time are shown in the bottom panel, with the audio signal divided into one-second intervals for the purpose of classification. Throughout the duration of the recording, the classifier identifies the presence of drones but does not attribute any segments to the C3 class. This indicates that while the classifier can detect drones, it fails to discern the specific C3 category correctly.

3.3. Single augmentation techniques

3.3.1. Harmonic Distortions

The impact of harmonic distortions on the accuracy of drone classification was investigated by varying the level of distortion applied to the training data. The distortion level was adjusted from 0% (indicating no augmentation) to 63%, with several gradations in between. The experiment was performed two times, without and with setting an initial seed prior to the training process. The results of the seed run in Table 3 suggest that slight to moderate distortion levels, specifically around 7% to 14%, tend to enhance the classifier's accuracy, which is in line with the optimal range identified in the run without seed initialisation. This level of distortion likely simulates the diverse sound qualities produced by drones in different operational conditions, thus potentially improving the model's adaptability to real-world situations.

The data indicates that as the harmonic distortion level increases, there is an initial improvement in classification accuracy, peaking within the range of 7% to 14%. This reinforces the findings of the non-seeded run, suggesting that a moderate level of distortion indeed enhances the model's accuracy by emulating the sound variations encountered in practical drone operations. Beyond this optimal distortion range, however, the accuracy starts to wane, indicating that while a controlled amount of distortion can be beneficial by preparing the model for a variety of real-world acoustic conditions, too much distortion introduces noise that can confuse the classifier. This demonstrates the necessity for a carefully calibrated approach to the application of harmonic distortion, emphasizing the balance between augmentation and the preservation of classification integrity to ensure the model remains effective when deployed in real-world environments.

The examination of harmonic distortions and their influence on the classifier's performance in the outdoor experiment from Figure 3 did not reveal a marked improvement when applying the classifier with the optimal distortion range determined from the seeded run. These results, from the application of classifiers augmented with harmonic distortion to a real-world outdoor setting, suggest

 doi:10.20944/preprints202401.2114.v1

that while moderate distortions enhance accuracy in controlled environments, the transition to outdoor conditions presents additional complexities.

Table 3. Classification accuracy for individual drone categories C0 to C3 across different levels of **augmentation with harmonic distortion**, with the last two columns displaying the average (Mean) and standard deviation (Std) of the accuracies for all categories. Results that meet or exceed the 75th percentile threshold for their category are highlighted in green, indicating higher accuracy, while results at or below the 25th percentile are highlighted in orange, indicating lower accuracy. For the standard deviation (Std), this color scheme is reversed: lower values (indicating more consistent accuracy) are marked in green and higher values (indicating less consistency) in red.

Distortion-Level	C0	C1	C2	C3	Mean	Std
0%	95.0%	97.7%	98.0%	98.9%	97.4%	1.7%
7%	97.2%	94.1%	96.2%	98.5%	96.5%	1.9%
14%	95.9%	96.7%	95.3%	99.1%	96.8%	1.7%
21%	95.1%	91.5%	93.2%	98.9%	94.7%	3.2%
28%	93.3%	89.1%	94.4%	98.5%	93.8%	3.9%
35%	92.5%	92.9%	94.8%	97.9%	94.5%	2.5%
42%	88.7%	96.1%	94.1%	98.8%	94.4%	4.3%
49%	89.6%	89.9%	94.2%	97.7%	92.9%	3.9%
56%	87.6%	92.6%	95.0%	96.9%	93.0%	4.0%
63%	90.0%	86.8%	93.8%	92.1%	90.0%	3.0%

3.3.2. Environmental Noise

A rigorous investigation of the effects of environmental noise augmentation on training data was conducted. Levels of noise introduced varied from 0% to 72%, with the aim of determining the impact of different noise intensities on the classifier's accuracy. Results summarized in Table 4 indicate that the incorporation of noise generally results in a decrease in classification accuracy. This finding is consistent with the discussion in the Methods section, which focused on the selective application of augmentation to training data. It is based on the understanding that augmented data may not always accurately replicate real-world conditions [19].

Table 4. Classification accuracy for individual drone categories C0 to C3 across different levels of **environmental noise augmentation**, with the last two columns displaying the average (Mean) and standard deviation (Std) of the accuracies for all categories. Results that meet or exceed the 75th percentile threshold for their category are highlighted in green, indicating higher accuracy, while results at or below the 25th percentile are highlighted in orange, indicating lower accuracy. For the standard deviation (Std), the color scheme is reversed: lower values are marked in green to indicate consistency, and higher values in red to indicate less consistency.

maxNoise	C0	C1	C2	C3	Mean	Std
0%	95.3%	96.9%	97.6%	98.0%	97.0%	1.2%
8%	96.7%	96.2%	98.3%	98.8%	97.5%	1.2%
16%	93.3%	92.1%	97.4%	98.6%	95.4%	3.1%
24%	95.4%	95.9%	97.3%	98.8%	96.9%	1.5%
32%	78.3%	93.6%	96.9%	99.5%	92.1%	9.5%
40%	64.7%	92.4%	96.3%	98.7%	88.0%	15.8%
48%	76.2%	88.0%	97.5%	81.0%	85.7%	9.3%
56%	76.1%	90.5%	90.9%	99.4%	89.2%	9.7%
64%	67.1%	90.3%	95.9%	93.4%	86.7%	13.2%
72%	50.9%	89.2%	96.9%	81.4%	79.6%	20.2%

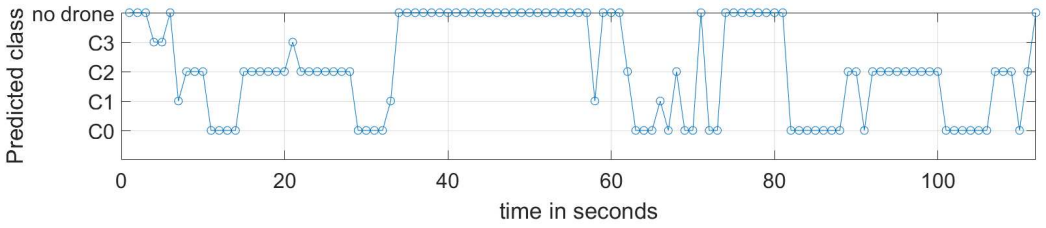
The examination of noise levels revealed that classifier performance remains relatively stable at noise levels up to 24%. This stability suggests that a controlled amount of noise could be beneficial, potentially contributing to the robustness of the classifier in real-world operational conditions where

 doi:10.20944/preprints202401.2114.v1

some background noise is inevitable. However, beyond a 24% noise level, a noticeable decline in accuracy is observed, with significant deterioration occurring at levels of 32% and above.

The application of noise-augmented classifiers to real-world data has unveiled intriguing insights, as displayed in Figure 4. The introduction of 24% environmental noise resulted in the drone being correctly classified as a C3 class in 3.1% of instances (which is more than 0% compared to the case without augmentation in section 3.2). With a 32% noise augmentation, this rate improved dramatically to 27.2% correct classifications during the recording period.

(a) 24 percent noise augmentation



(b) 32 percent noise augmentation

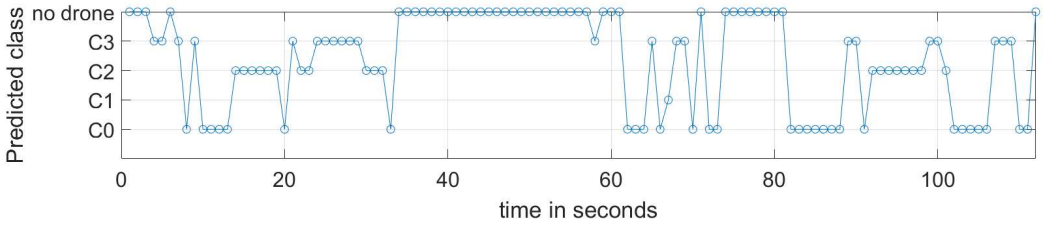


Figure 4. Classification results over time for a C3 drone showing the classifiers predictions, trained with different degrees of no noise amplitude.

The increase in classification accuracy observed at higher noise levels appears counterintuitive, given the typical perception of noise as harmful to signal clarity. Nonetheless, it is imperative to interpret these results in the context of the validation methodology employed. As outlined in the methods section, augmentation was exclusively applied to the training data. This strategy aimed to maintain the realism of the validation set and to avoid the introduction of potential artifacts that could bias the results [19]. Such a careful approach was adopted to ensure that the evaluation of the classifiers' performance was conducted using unaltered, real-world data, which was anticipated to yield a more precise assessment of their effectiveness.

3.3.3. Pitching

The evaluation of pitch augmentation's effect on the classifier's performance involved adjusting the maximum pitch change parameter, referred to as maxPitch. For clarity, maxPitch is defined as the boundary within which each audio chunk's pitch could be randomly altered, ranging from -maxPitch to +maxPitch semitones, where 0 signifies no augmentation. This parameter was varied from 0 to 0.9 semitones. The outcomes of this pitch variation are documented in Table 5, where it is shown that changes in pitch produce a range of effects on the classifier's performance, with the impact of these changes being mostly marginal when compared to the baseline. The exploration included an analysis of the pitch's influence on the recognition accuracy across different drone categories.

 doi:10.20944/preprints202401.2114.v1

Table 5. Classification accuracy for individual drone categories C0 to C3 across different levels of **pitch augmentation**, with the last two columns displaying the mean (Mean) and standard deviation (Std) of the accuracies for all categories. Results that meet or exceed the 75th percentile threshold for their category are highlighted in green, indicating higher accuracy, while results at or below the 25th percentile are highlighted in orange, indicating lower accuracy. For the standard deviation (Std), this color scheme is reversed: lower values (indicating more consistent accuracy) are marked in green and higher values (indicating less consistency) in red.

maxPitch	C0	C1	C2	C3	Mean	Std
0	96.6%	98.2%	97.6%	99.2%Ï	97.9%	1.1%
0.2	93.7%	97.0%	98.2%	99.0%	97.0%	2.3%
0.4	95.4%	96.9%	97.3%	99.4%	97.3%	1.7%
0.6	94.2%	95.2%	95.9%	97.6%	95.7%	1.4%
0.8	92.3%	97.2%	98.1%	98.9%	96.6%	3.0%
1.1	94.1%	94.2%	94.9%	98.0%	95.3%	1.8%
1.4	92.3%	95.1%	96.1%	98.8%	95.6%	2.7%
1.7	93.0%	94.1%	97.0%	97.8%	95.5%	2.3%
2.1	94.6%	96.8%	95.5%	97.4%	96.1%	1.3%
2.5	92.9%	93.0%	96.8%	98.4%	95.3%	2.8%

The exploration of pitch augmentation's influence on the classifier's performance has indicated that subtle pitch modifications, up to +/- 0.4 semitones, do not substantially decrease the model's precision. This fine-tuning of pitch may indeed bolster the classifier's resilience, maintaining or possibly improving its effectiveness, as illustrated by the findings shown in Figure 5. The utilization of the classifier, with a pitch augmentation around +/- 0.4 semitones, on drone flight data as discussed in Section 3.2, achieved the correct drone class identification in four out of 71 cases. While this number might not seem impressive at first glance, it marks a significant improvement over the non-augmented case, especially considering that there was a notable increase in classifications in the C2 category instead of C0. This refinement is advantageous as it is preferable to have a misclassification within one class difference rather than three, highlighting an improvement in the classifier's discriminative power. Such enhancements in classification accuracy, even if modest, underline the potential of minimal pitch augmentation to slightly enhance robustness and improve the model's performance in real-world scenarios. Larger pitch changes, however, are observed to diminish accuracy, reinforcing the concept that while considerable alterations may be detrimental, moderate pitch adjustments are likely to be innocuous or advantageous. This insight is crucial for fine-tuning pitch augmentation strategies to improve the accuracy and robustness of drone sound identification in practical applications.

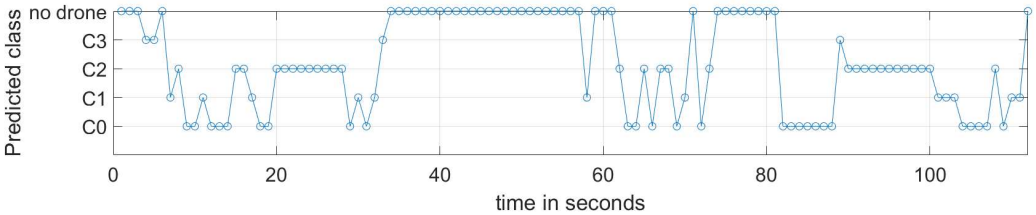


Figure 5. Classification results over time, showing the classifier's predictions with an Augmentation with a pitching of about +/- 0.4 semitones.

3.3.4. Echos

The impact of introducing a delay in the audio signal on the classifier's performance was evaluated. Delays ranged from 15 ms to 27 ms, with varying maximum amplitudes from 30% to 90%. The goal of this augmentation was to replicate real-world acoustic phenomena, such as echoes, commonly encountered in different environments, including ground reflections. The summarized results can be found in Table 6.

 doi:10.20944/preprints202401.2114.v1

Table 6. Classification accuracy for individual drone categories C0 to C3 across different levels of **delay augmentation**, with the last two columns displaying the mean (Mean) and standard deviation (Std) of the accuracies for all categories. Results that meet or exceed the 75th percentile threshold for their category are highlighted in green, indicating higher accuracy, while results at or below the 25th percentile are highlighted in orange, indicating lower accuracy. For the standard deviation (Std), this color scheme is reversed: lower values (indicating more consistent accuracy) are marked in green and higher values (indicating less consistency) in red.

Delay Duration	Delay Amplitude	C0	C1	C2	C3	Mean	Std
15 ms	30%	97.9%	96.2%	97.3%	98.8%	97.6%	1.1%
15 ms	50%	96.7%	95.4%	97.6%	98.8%	97.1%	1.4%
15 ms	70%	94.7%	96.9%	96.8%	98.5%	96.7%	1.6%
15 ms	90%	94.6%	96.1%	97.3%	98.3%	96.6%	1.6%
18 ms	30%	94.1%	94.4%	96.6%	99.1%	96.1%	2.3%
18 ms	50%	93.3%	97.0%	97.6%	98.7%	96.7%	2.3%
18 ms	70%	88.4%	97.4%	97.6%	98.9%	95.6%	4.8%
18 ms	90%	94.9%	97.3%	96.9%	98.7%	97.0%	1.6%
21 ms	30%	96.6%	96.7%	96.1%	97.7%	96.8%	0.7%
21 ms	50%	95.5%	97.2%	97.8%	99.1%	97.4%	1.5%
21 ms	70%	95.0%	96.2%	97.1%	98.3%	96.7%	1.4%
21 ms	90%	92.1%	95.3%	97.9%	99.5%	96.2%	3.2%
24 ms	30%	94.3%	97.6%	97.4%	98.6%	97.0%	1.9%
24 ms	50%	95.2%	95.7%	97.1%	98.6%	96.7%	1.5%
24 ms	70%	92.5%	96.9%	95.9%	98.7%	96.0%	2.6%
24 ms	90%	93.7%	94.1%	96.6%	98.8%	95.8%	2.4%
27 ms	30%	93.5%	97.5%	97.0%	98.5%	96.8%	2.2%
27 ms	50%	93.3%	95.8%	96.9%	98.8%	96.2%	2.3%
27 ms	70%	93.7%	97.0%	96.2%	98.3%	96.3%	1.9%
27 ms	90%	94.6%	94.5%	96.8%	99.2%	96.3%	2.2%
random		92.9%	97.4%	96.7%	99.3%	96.6%	2.7%

The table does not reveal a specific trend. Furthermore, the differences in accuracy for the various parameters are minimal and may be attributable to noise. Unlike previous augmentation techniques, this evaluation requires a reference to the specific measurement situation. Delay primarily simulates ground reflection, which heavily relies on the coordinates of the microphone and drone, as well as the surface's reflectivity. In realistic scenarios, this leads to time differences between the direct signal and reflection (delay) ranging from 0 (microphone at ground level) to 30 ms, and amplitude differences based on the reflection coefficient ranging from nearly 0 (on highly absortive surfaces like grass) to 1 (concrete ground). Consequently, it is essential to encompass as much variability as possible in the delay augmentation.

When applying various delay classifiers to the previously discussed example of the C3 drone, as shown in Figure 6 (based on 21 ms delay and 30% amplitude), up to 8.2% of drone events are correctly identified as C3 drones. However, the accuracy of recognition significantly depends on the drone's location and the associated delay between its direct signal and its echo. For echo augmentation, it is imperative that the delay and amplitude parameters cover the entire range of realistic reflections as comprehensively as possible.

Therefore, an additional classifier was created in which the values for the delay duration between 0 and 30 ms and the delay amplitude based on a reflection coefficient between 0 and 1 were randomly varied in the training data for each audio chunk. This is emphasized in the table by the indication 'random' in the columns 'Delay Duration' and 'Delay Amplitude'. The performance of this classifier compared to the others, as measured by the accuracy results, does not appear to offer any advantage.

 doi:10.20944/preprints202401.2114.v1

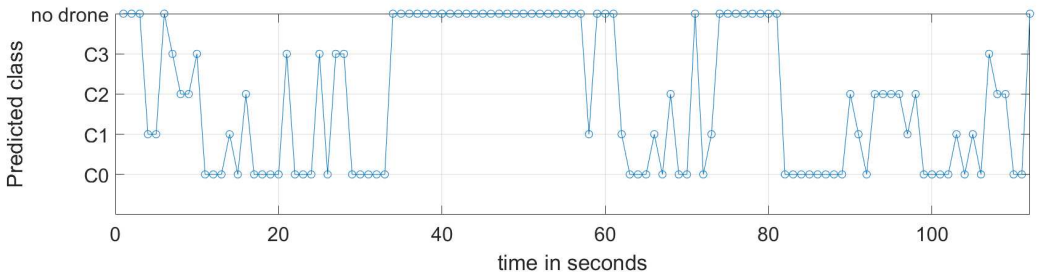


Figure 6. Classification results over time, showing the classifier's predictions with an Augmentation with a an echo of about 21 ms an 30% amplitude.

However, when the classifiers with random delay augmentation are applied to the sample measurement (see Figure 7), in 10.7% of all cases where the binary classifier has detected a drone, the correct drone class is subsequently identified, which was the best improvement of all delay augmenters. This underlines the importance of comprehensive coverage of realistic reflectance parameters in delay augmentation.

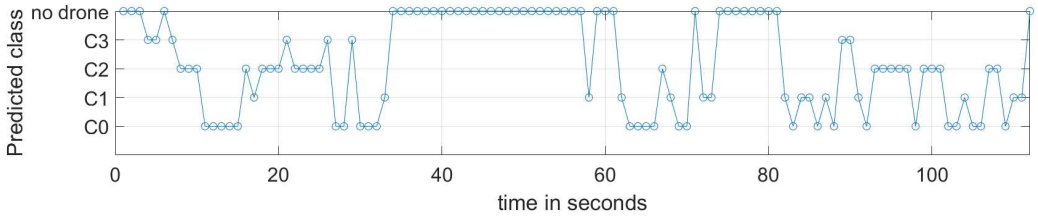


Figure 7. Classification results over time, showing the classifier's predictions with an Augmentation with arbitry deleays (delay time varying between 0 and 30 ms; reflection coefficient varying between 0 and 1).

3.4. Combined augmentation techniques

The study's focus on real-world performance assessment involved analyzing various classifiers, each employing a combination of augmentation techniques with specified parameter settings. The key parameters varied were maxPitch (0.36 to 0.48), maxNoise (23% to 41%), and distLevel (16% to 24%). Additionally, delay parameters were set randomly for each data chunk (delay durations of 0 to 30 milliseconds and amplitude variations from 0 to 1). Recognizing the limitations of solely relying on confusion matrices for accuracy evaluation, as discussed earlier, a different approach was taken for this analysis. For each drone class, a specific drone recording from our database, not part of the anechoic chamber measurements, was selected. The percentage of correct drone class identification was then determined over the duration of each drone flight. The following insights were gleaned from this analysis:

1. Effectiveness in Lower Drone Categories (C0 and C1):
    - The augmentation techniques, particularly variations in maxPitch, maxNoise, and distLevel, showed promising results in the lower drone categories (C0 and C1).
    - Specific augmentations led to a noticeable improvement in accurately classifying these categories, including when considering the neighboring class hits.
    - The classifiers generally demonstrated an acceptable level of accuracy for C0 and C1 categories, suggesting that the chosen augmentation methods were effective for smaller drones.
2. Challenges in Higher Drone Categories (C2 and C3):

 doi:10.20944/preprints202401.2114.v1

- In contrast, the classification of larger drones (C2, and especially C3) was less successful. The same augmentation techniques that benefited lower categories did not translate well to these higher categories.
- The accuracy, even when considering hits in the neighboring classes, was not deemed acceptable for C2 and was particularly lacking for C3.
- This indicates a need for further refinement in the augmentation approach or possibly the development of distinct strategies tailored to larger drone categories.

The results from this approach showed a general tendency of the classifiers to categorize drone recordings as C0 or C1. This trend highlights a potential bias in the classifiers towards lower drone categories and emphasizes the need for further exploration and refinement in the classification of higher drone categories.

4. Discussion

4.1. Interpretation of Findings

This study focused on developing an advanced drone classification system based on acoustic signatures, enhanced by sophisticated data augmentation techniques. Key interpretations of the findings are as follows:

- **Influence of Random Processes in Model Training:** In this study, the influence of random processes during the training process, such as the initialization of weights, on the results has been importantly highlighted. It has been strongly observed that the setting of a fixed seed for random number generation significantly reduces variability in ML model training outcomes. This contributes to more reliable and reproducible results, emphasizing the necessity of managing random initialization effects in ML experiments. Consequently, the employment of fixed seeds is advocated as a best practice for achieving consistency in ML model performance. It is recommended for future studies to utilize fixed seeds for random number generators, standardizing weight initialization and the selection of mini-batches. This approach ensures more consistent outcomes and interpretations in ML research.
- **Impact of Data Augmentation:** The results demonstrate that different data augmentation techniques – specifically pitch shifting, time delay, harmonic distortion, and ambient noise – improve the classifier's performance adapted to real situations. The study underscores that each augmentation method affects specific aspects of drone detection and thus needs to be carefully tailored to the dataset in question.
- **Harmonic Distortion:** Introducing harmonic distortions helped the system simulate the effects of sound waves traveling through different mediums, improving accuracy in complex acoustic environments.
- **Inclusion of Ambient Noise:** Adding ambient noises to the training data helped prepare the model to distinguish drone sounds against the backdrop of everyday noises. However, it was found that too much noise can impair the classifier's performance.
- **Pitch Shifting:** Adjusting the pitch within a certain range enabled the system to be more flexible in responding to variations in drone motor sounds. The results indicate that slight adjustments in pitch can enhance detection accuracy, noting that excessive changes can be counterproductive.
- **Time Delay and Echo Effects:** Time delays were introduced in the audio signal to mimic echo effects in various environments, resulting in increased adaptability of the model to different acoustic conditions. Recent experiments with random delay augmentation have emphasized the complexity of simulating real-world echo conditions. Although the classifier's accuracy did not notably improve with this method, its application in real-world scenarios, characterized by widely varying echo characteristics, demonstrated an increased ability to correctly identify drone classes. This indicates the importance of incorporating a broad range of delay parameters to capture the variety of possible real-world echo conditions.

 doi:10.20944/preprints202401.2114.v1

- **Combined Application of Techniques:** The combination of these augmentation techniques yielded mixed results. Generally, all drone recordings were disproportionately classified into drone classes C0 and C1. A clear tendency on how the parameters for the various techniques should be chosen for simultaneous augmentation was not discernible.

Overall, this study demonstrates the potential of acoustic data augmentation techniques to enhance the efficiency of ML systems for drone detection. The results provide valuable insights into the development of effective UAV detection systems that can be utilized in various security and management contexts. The exploration of random delay and amplitude variation particularly underscores the need to consider a wide spectrum of realistic acoustic conditions for more accurate drone classification.

4.2. Theoretical and practical implications

Significant contributions to the advancement of UAV detection technology have been made through this research. The inclusion of 44 different drone models in an extensive database establishes a robust foundation for the development and training of ML algorithms in drone detection. The capability of the classifier to detect and categorize drones into EU-defined classes C0 to C3 is underscored, emphasizing the practical applicability of the study. Additionally, the potential versatility of the classifier across a variety of operational environments is highlighted.

It is essential to consider that training with audio data from outdoor recordings inherently captures elements of the specific scene, such as reflections, ambient noise, and other environmental factors. Even recordings made in complete stillness on an open, flat field or parking lot will at least encompass ground reflections, which vary depending on the ground's characteristics and the relative positions of the drone and microphone. Conversely, free-field data recorded in an anechoic chamber can be augmented to represent arbitrary scenarios. Such augmentation allows the creation of a controlled yet diverse set of training data that can better prepare the model for the unpredictability of real-world acoustic environments.

The flexibility in augmentation both theoretically and practically has significant implications. From a theoretical perspective, it is challenged and refined in understanding the influence of environmental factors on the processing and classification of acoustic signals in ML models. From a practical standpoint, a methodology is provided to enhance the adaptability of detection systems, ensuring their effectiveness across a broad spectrum of real-world conditions. Systematic augmentation of clean, free-field data allows for the simulation of various operational scenarios, effectively expanding the classifier's exposure to potential real-world acoustic signatures.

4.3. Limitations and Challenges

The work on UAV detection using ML presented in this paper is characterized by specific limitations and challenges, evident both in the experimental results and within the context of the cited literature. A principal limitation, as highlighted in [12], is the inherent variability of ML model performances, even with the use of identical parameter settings. This variability is attributed to random processes in initialization and learning algorithms, emphasizing the necessity for meticulous validation and robust design in model development to achieve consistent and reliable classification results.

A significant limitation is presented by the close relationship between the validation and training data sets. Despite the decision, based on guidance from [19], not to apply augmentation techniques to the validation data, the fact that this data is derived from the same set of measurements as the training data introduces a risk of overfitting. This strategy aligns with the understanding that augmented data may not accurately represent realistic scenarios or could introduce alteration artifacts [19]. However, it could result in classifiers being unduly tuned to the characteristics of the training environment. This may lead to an overestimation of their efficacy in real-world applications due to their familiarity with the data. This aspect is crucial, as it could compromise the classifiers' ability to generalize to new, unseen data, which is a vital criterion for real-world applications.

 doi:10.20944/preprints202401.2114.v1

Furthermore, the adaptability of the classifier to real-world scenarios, characterized by acoustic diversity, remains a challenge. Despite the extensive database [10] and sophisticated augmentation techniques, the accurate classification of drone sounds in dynamic environments proves difficult. Studies such as [5–8] confirm the complexity of distinguishing drones against various background noises and conditions.

The significance of optimally tuning data augmentation techniques to maximize the classifier's accuracy and efficiency is also a notable challenge. Techniques like pitch shifting and time delay, as mentioned by [19], are acknowledged for enhancing model adaptability. However, their integration must be balanced carefully to prevent overfitting or performance degradation. It has been indicated that the relationship between noise levels and classification performance is complex and non-linear. Certain noise levels might unexpectedly improve the classifier's discriminative capabilities, underscoring the importance of noise management in training data for the model's applicability in real-world scenarios.

The focus of the study was on four specific augmentation techniques: pitch shifting, adding delay, introducing harmonic distortions, and incorporating environmental noise. The field of data augmentation in ML, especially concerning acoustic signal processing, is extensive and continuously developing. Future research could investigate additional methods and parameters. These might include manipulating spectrograms [13], altering volume, adding reverberation effects, and creating synthetic data using methods like Generative Adversarial Networks (GANs), as discussed in [6].

An additional aspect that must be contemplated is the possibility that classification into distinct drone classes based solely on acoustic signatures may not be entirely feasible. The primary source of noise in drones typically emanates from their rotors. Changes in rotor types can lead to noticeable alterations in the drone's acoustic footprint. For instance, Figure 8 (top) displays the Fast Fourier Transforms (FFT) of the same hovering C2 drone (a Phantom 4 RTK) recorded in an anechoic chamber from our measurement campaign, pre and post the rotor change. The FFTs show principally similar patterns, yet notable differences are observed with Rotor 2, including a slight shift of harmonics to lower frequencies and a significant decrease at higher frequencies. Conversely, comparing the Phantom 4 RTK with Rotor 1 to a different C2 drone (DJI Mavic 3e) reveals a much more pronounced change in the acoustic fingerprint. These observations suggest that distinguishing between different types of drones, irrespective of the rotor, is feasible. However, the possibility of generalizing acoustic fingerprints based on drone classes remains an open question and should be the focus of future research.

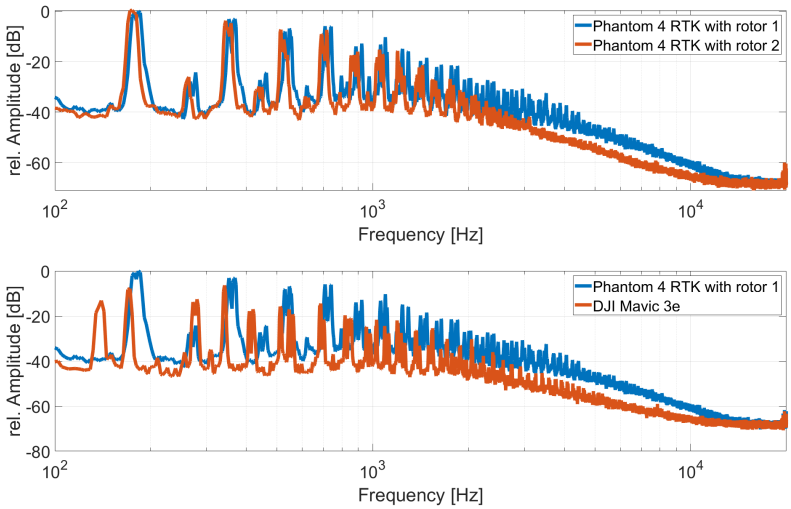


Figure 8. FFT Analysis of Drone Acoustic Signatures: Effects of Rotor Change and Comparison Between Different Drone Models

 doi:10.20944/preprints202401.2114.v1

4.4. Future Research

As indicated by the recent spectral analyses, the logical next step is a systematic examination of the features specific to each drone class. A multitude of potential features can be explored. These could include time-domain features such as zero crossing rate or short-time energy, frequency-domain features like chromagrams, spectral roll-off, or linear predictive coding (LPC) coefficients, cepstral-domain features such as mel-frequency cepstral coefficients (MFCCs), linear prediction cepstral coefficients, or Greenwood function cepstral coefficients, as well as image-based features, among others. A comprehensive overview of various features can be found in [20].

This exploratory approach would involve a detailed investigation into the distinctive acoustic signatures of different drone classes. By analyzing a diverse array of features, it may be possible to uncover unique patterns and characteristics that differentiate one class from another. This research would not only contribute to a deeper understanding of drone acoustics but also significantly enhance the precision and reliability of drone classification systems.

Building on this foundational research, the use of outdoor measurement campaign data as validation data presents a promising approach for assessing the performance of classifiers in real-world drone noise classification scenarios. This method could lead to enhanced generalizability and reliability of models in complex environments, though it requires meticulous preparation and analysis of the measurement data.

The challenge lies in preparing the recordings to provide representative and usable validation data. A critical step would be identifying and filtering out segments without drone noise, such as periods before takeoff, after landing, and when the drone is too distant. For this purpose, the already developed 'Drone' vs. 'No Drone' classifier could be utilized. Despite being trained solely on unaugmented free-field data, this classifier has proven to be effective and robust. Its use in preprocessing the outdoor measurement data could be a practical solution for identifying relevant segments for validation.

The advantages of this approach for future research include:

1. **Realistic Validation:** Using outdoor measurement data allows for a more realistic and meaningful assessment of classifiers, helping to evaluate the models' ability to operate accurately under various real conditions.
2. **Data Quality Improvement:** By selectively filtering out irrelevant sections, the validation data can gain in quality, leading to more accurate assessments of model performance.
3. **Efficiency Enhancement:** The use of the 'Drone' vs. 'No Drone' classifier for preprocessing can significantly speed up and simplify the data preparation process.
4. **Insight for Model Improvements:** The results from the validation can provide insights into necessary model improvements, particularly in terms of robustness against environmental noises and other variable factors.

However, this approach also brings challenges. Preparing the measurement data requires careful analysis and potentially the development of new methods for data segmentation and classification. Additionally, it is essential to ensure that the validation data covers a sufficient variety of scenarios and background noises to allow for a comprehensive assessment of the classifiers.

Integrating this method into future research could lead to a significant increase in the accuracy and reliability of drone detection systems. It provides a solid foundation for the development of more advanced models capable of operating effectively in a variety of environments. The database should be expanded to include more outdoor drone recordings for drone classes C0 and C1. This expansion would provide a more comprehensive dataset for training and validating classifiers, contributing to the development of more accurate and reliable drone detection systems capable of distinguishing between various drone classes in diverse outdoor environments. In this context, exploring other aspects, such as investigating different neural network architectures or experimenting with new acoustic features, would continue to play a role. This comprehensive approach would ultimately help to push the

 doi:10.20944/preprints202401.2114.v1

boundaries of current technologies and develop more robust, adaptable solutions for the challenges of drone detection.

In addition to the fundamental revision of the methodological approach of using the outdoor data as validation data during the training process, future evaluations on the relationship between the classification results and the drone's distance from the microphones would be beneficial. This analysis is crucial to determine the operational limits of the classifiers in terms of detection range. It involves assessing how the drone's volume, relative to surrounding environmental noises, impacts classification accuracy. This investigation would provide key insights into the minimum distances at which drones of various classes can be effectively detected and classified. Additionally, it would explore the interplay between the loudness of drone sounds and ambient noises, which is vital in dynamic and unpredictable acoustic environments.

In this context, it is also important to examine the influence of loudness on drone classification in more detail. In this study, all audio segments were normalized to 90 percent of their maximum amplitude in order to minimize the influence of different volume levels. This normalization serves to ensure the consistency of the input data for the classification process. However, information contained in the volume of the drone sounds, especially with regard to the distance of the drone to the microphone, could be of great value. Future research could therefore include more detailed investigations into the effects of different loudness levels on classification results.

A possible extension of this investigation could be the use of compression techniques. The use of compression could achieve a balance between reducing loudness differences and maintaining sufficient dynamic range for classification. This could be particularly beneficial in environments with highly varying background noise, as appropriate compression could help to emphasize the characteristic features of drone sounds even at different distances and background volumes.

An additional aspect of future research that warrants exploration is the impact of multiple reflections on drone sound classification. In this study, we primarily focused on simple delays, analogous to ground reflections, which are undoubtedly crucial in drone sound analysis. However, in real-world scenarios, acoustic signals often undergo multiple reflections, leading to complex reverberation patterns. These multi-path effects, akin to reverb in audio processing, can significantly alter the perceived drone sound, especially in urban environments or indoors. Investigating the influence of such multiple delays on the classification system is therefore essential. This investigation would involve simulating various reflective conditions, from simple two-surface reflections to more complex, hall-like reverberations, and assessing their impact on the classifier's ability to accurately identify and categorize drone sounds.

5. Conclusions

In the study "Sound of Surveillance: Enhancing ML-Driven Drone Detection with Advanced Acoustic Augmentation," a comprehensive exploration is presented into the application of advanced acoustic data augmentation techniques for improving the performance of ML systems in drone detection. The key conclusions drawn from the research are:

- **Effectiveness of Data Augmentation:** Various data augmentation techniques, such as pitch shifting, time delay, harmonic distortion, and ambient noise incorporation, have been demonstrated to significantly enhance the classifier's accuracy. These techniques have been shown to enable the system to adapt to diverse acoustic environments, effectively identifying and categorizing drone sounds amidst a variety of background noises.
- **Optimization of Augmentation Techniques:** The study reveals that each augmentation method impacts specific aspects of drone sound detection. For instance, moderate levels of pitch shifting and harmonic distortion were found to be most effective, while excessive application of these techniques could lead to reduced performance or counterproductive results. Additionally, introducing time delays and ambient noises at controlled levels improved the model's robustness and adaptability.

 doi:10.20944/preprints202401.2114.v1

- **Classifier Performance and Reproducibility:** The research highlighted the critical role of random processes in ML model training. Variability in the performance of classifiers, even under identical parameter settings, underscores the importance of ensuring consistent initialization of initial weights and the selection of mini-batches. Future research should prioritize standardizing these aspects to achieve more reliable and reproducible outcomes.
- **Practical Applicability and Future Directions:** While the current ML-based classifier demonstrates significant potential in security and airspace management, complying with EU drone categorization regulations, further refinement is required for optimal performance in classifying drone categories. General drone detection using ML has proven effective, yet precise categorization of drones into specific classes as per EU standards demands additional research. Future studies should focus on exploring advanced optimization algorithms and experimenting with diverse parameter combinations. This exploration will be critical for enhancing the accuracy of drone noise classification systems, particularly in accurately identifying and classifying drones into distinct regulatory categories. Continued research in this direction will not only improve the reliability of drone detection systems but also ensure their compliance with evolving regulatory frameworks, thereby bolstering their practical applicability in various real-world scenarios.
- **Contribution to UAV Detection Technology:** Significant contributions have been made to the field of UAV detection technology through this research. The establishment of a comprehensive database encompassing 44 different drone models provides a solid foundation for the continued development and training of ML algorithms in this domain. The demonstrated capability to classify drones into distinct categories (C0 to C3) in accordance with EU regulations underlines the practical applicability and relevance of the system in meeting both current and emerging requirements in drone security.

In summary, this study offers valuable insights into the development of effective UAV detection and classification systems, leveraging sophisticated acoustic data augmentation techniques. It lays a foundation for future research and advancements in this field, aimed at enhancing security and management capabilities in response to the growing use of drones.

Author Contributions: This research was conducted with the primary contributions from the sole author, Sebastian Kümmritz (SK). The conceptualization, methodology, software development, validation, formal analysis, investigation, resource provision, data curation, original draft preparation, manuscript review and editing, visualization, project administration, and funding acquisition were all carried out by SK. The author also read and agreed to the published version of the manuscript.

Special acknowledgment is given to Ernst Swanepoel (swanepoel@h2think.org) for creating Figure 1 and his significant role in conducting the measurement campaigns. Furthermore, gratitude is extended to Lothar Paul from H2 Think for his essential work in establishing the drone sound database, as referenced in [10]. While their contributions were invaluable to the project, it was decided that they would not be included in the authorship of this paper, in line with the authorship guidelines that limit inclusion to those who have contributed substantially to the work reported (as per CRediT taxonomy).

Funding: This research, part of the project "AuDroK" with grant number 19F1131A, was funded by the Federal Ministry of Digital and Transport of the Federal Republic of Germany under the mFund mechanism (https://bmdv.bund.de/DE/Themen/Digitales/mFund/Ueberblick/ueberblick.html), covering 80% of the project costs. The remainder was financed from internal resources.

Data Availability Statement: This study has developed a comprehensive database of classified acoustic drone recordings, in alignment with EU drone regulations [11], accessible at [10] https://mobilithek.info/offers/605778370199691264. Additionally, the source code for the algorithms used in this research is available and can be referenced at [10].

Acknowledgments: In this section you can acknowledge any support given which is not covered by the author contribution or funding sections. This may include administrative and technical support, or donations in kind (e.g., materials used for experiments).

Conflicts of Interest: "The authors declare no conflicts of interest." "The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results".

 doi:10.20944/preprints202401.2114.v1

Abbreviations

The following abbreviations are used in this manuscript:

EU	European Union
FFT	Fast Fourier Transform
MFCC	Mel-Frequency Cepstral Coefficients
ML	Machine Learning
UAV	Unmanned Aerial Vehicles

Appendix A

Appendix A.1 Source code listings

Listing 1: Pitching Function 'applyCustomPitching'

```
\small

function augAudio = applyCustomPitching(audioIn, semiTone, Fs)
  % Generate a random pitch change within the [-semiTone, semiTone] range
  pitchShift = (rand()*2 -1)*semiTone;

  % Conversion of pitch shift to frequency ratio
  freqRatio = 2^(pitchShift / 12);

  % Check whether a reduction in the sampling rate is necessary
  originalFs = Fs;
  if Fs > 48000
    % Temporary reduction of the sampling rate for
    % processing
    Fs = 48000;
    audioIn = resample(audioIn, Fs, originalFs);
  end

  % Applying the pitch shift
  augAudio = pitchShiftWithSameSpeed(audioIn, freqRatio, Fs);

  % Return to the original sampling rate if necessary
  if originalFs > 48000
    augAudio = resample(augAudio, originalFs, Fs);
  end
end

% is required for applyCustomPitching
function shiftedAudio = pitchShiftWithSameSpeed(audioIn, freqRatio, Fs)
  % Adjusting the sampling rates to integers for the resample function
  segmentLength = 1e5; % Length of each segment
  numSegments = ceil(length(audioIn) / segmentLength);
  shiftedAudio = [];

  newFs = round(Fs * freqRatio);

  for i = 1:numSegments
    segmentStart = (i - 1) * segmentLength + 1;
    segmentEnd = min(i * segmentLength, length(audioIn));
    segment = audioIn(segmentStart:segmentEnd);

    % Applying the pitch shift to the current segment
    try
      shiftedSegment = resample(segment, newFs, Fs);
    catch
      shiftedSegment = segment;
```

 doi:10.20944/preprints202401.2114.v1


```
      disp('Error with resampling')
    end

    % Adding the edited segment to the output signal
    shiftedAudio = [shiftedAudio; shiftedSegment];
  end
end
```

Listing 2: Delay Function 'applyDelay'

```

function augAudio = applyDelay(audioIn, maxDelay, maxAmp, sampleRate)

  % Check whether the delay is too long
  if maxDelay > length(audioIn)/sampleRate
    error('Maximum delay exceeds the signal length.');
  end

  if maxDelay == -1
    randomDelay = rand(1)*30e-3;
  else
    % Delay is varied by +/- 10
    randomDelay = maxDelay*(1+0.2*(rand()-0.5));
  end
  if maxAmp == -1
    randomAmp = rand(1);
  else
    % Amplitude is varied by +/- 10
    randomAmp   = maxAmp*(1+0.2*(rand()-0.5));
  end

  % Convert delay to samples
  delaySamples = round(randomDelay * sampleRate);

  % Generate delay signal
  delaySignal = [zeros(delaySamples, 1);
  audioIn(1:end-delaySamples)];

  % Multiply delay signal with random amplitude
  delaySignal = delaySignal * randomAmp;

  % Create output signal by superimposing
  try
    augAudio = audioIn + [zeros(delaySamples, 1);
    delaySignal(1:end-delaySamples)];
  catch
    disp('An error has occurred');
  end

  % Clipping test
  if max(abs(augAudio)) > 1
    augAudio = 0.9 * augAudio / max(abs(augAudio));
  end
end
```

Listing 3: Function 'applyEnvironmentalNoise' for adding environmental noise to the drone sounds

```

function audioOut = applyEnvironmentalNoise(audioIn, Fs_sig, maxAmp)
  % Folder with arbitrary non-drone noise (e.g. from youtube)
  folderPath = 'D:\Environmental Noise';
  % Get a list of all files in the folder
  files = dir(fullfile(folderPath, '*.wav'));

```

 doi:10.20944/preprints202401.2114.v1


```
    % If no file is found display a message and return
    if isempty(files)
      disp('No files found');
      audioOut = audioIn;
      return;
    end

    if maxAmp < 0 || maxAmp > 1
      disp('maxAmp must be between 0 and 1');
      audioOut = audioIn;
      return;
    end

    % Generate a random index into the files array
    randIndex = randi(length(files));

    % Select a random file
    randFile = files(randIndex).name;

    % Construct full file path
    fullFilePath = fullfile(folderPath, randFile);

    % Load the envirnonmental audio file
    [audioEnv, Fs_env] = audioread(fullFilePath);
    if Fs_env ~= Fs_sig
      % Resample signal
      audioEnv = resample(audioIn, Fs_sig, Fs_env);
    end

    % Take an arbitrary section with the same length
    % of the input file
    sInput = length(audioIn);
    sEnvir = length(audioEnv);

    if (sEnvir(1) - sInput(1)) > 0
      ind = randi(sEnvir(1) - sInput(1));
      audioEnv = audioEnv(ind:(ind+sInput-1));

      % Calculation of the rms of the input signal
      rms_sig = rms(audioIn);

      % Adjustment of the amplitude of the ambient noise
      rms_env = rms(audioEnv);
      scaled_audioEnv = (rms_sig / rms_env) * maxAmp * audioEnv;

      % Create mixed signal
      if sum(size(audioIn) == size(scaled_audioEnv)) == 0
        audioOut = audioIn + scaled_audioEnv';
      else
        audioOut = audioIn + scaled_audioEnv;
      end

      % Scaling if the signal is clipped
      if max(audioOut) > 0.9
        audioOut = 0.9* audioOut/max(audioOut);
      end
    else
      audioOut = audioIn;
    end
  end
```

Listing 4: Function 'applyHarmonicDistortion' for adding harmonic distortions to the drone sounds

```
    function augAudio = applyHarmonicDistortion(audioIn, distortionLevel, ...
```

```
                                                sampleRate)
    % Check whether the degree of distortion is valid
    if distortionLevel < 0 || distortionLevel > 1
      error('Distortion level must be between 0 and 1.');
    end

    % Add harmonic distortion through non-linear function
    augAudio = audioIn - distortionLevel * sin(2 * pi * (500/sampleRate) * ...
               (1:length(audioIn))') .* audioIn;

    % Clipping check and correction if necessary
    if max(abs(augAudio)) > 1
      augAudio = augAudio / max(abs(augAudio));
    end
  end
```

References

1. Gatwick Airport drone attack: Police have 'no lines of inquiry'. Available online: BBC News (accessed on 02 January 2024).
2. Knoedler, B.; Zemmari, R.; Koch, W. On the detection of small UAV using a GSM passive coherent location system. In Proceedings of the 17th International Radar Symp. (IRS), Krakow, Poland, Date of Conference (05-2016); https://ieeexplore.ieee.org/document/7497375/
3. Nguyen, P.; Ravindranatha, M.; Nguyen, A.; Han, R.; Vu, T. Investigating Cost-effective RF-based Detection of Drones. Proceedings of the 2nd Workshop on Micro Aerial Vehicle Networks, Systems, and Applications for Civilian Use, Singapore, Singapore, 26-06-2016, DOI: 10.1145/2935620.2935632
4. Shi, X.; Yang, C.; Xie, W.; Liang, C.; Shi, Z.; Chen, J. Anti-Drone System with Multiple Surveillance Technologies: Architecture, Implementation, and Challenges. *IEEE Commun. Mag.* **2018**, *56*, 68–74. DOI: 10.1109/MCOM.2018.1700430
5. Utebayeva, D.; Ilipbayeva, L.; Matson, E.T. Practical Study of Recurrent Neural Networks for Efficient Real-Time Drone Sound Detection: A Review. Drones **2022**, *7*, 26. DOI: 10.3390/drones7010026
6. Al-Emadi, S.; Al-Ali, A.; Al-Ali, A. Audio-Based Drone Detection and Identification Using Deep Learning Techniques with Dataset Enhancement through Generative Adversarial Networks. Sensors **2021**, *21*, 4953. DOI: 10.3390/s21154953
7. Dumitrescu, C.; Minea, M.; Costea, I.M.; Cosmin Chiva, I.; Semenescu, A. Development of an Acoustic System for UAV Detection. Sensors **2020**, *20*, 4953. DOI: 10.3390/s20174870
8. Jeon, S.; Shin, J.-W.; Lee, Y.-J.; Kim, W.-H.; Kwon, Y.-H.; Yang, H.-Y. Empirical Study of Drone Sound Detection in Real-Life Environment with Deep Neural Networks. *arXiv* **2017**. DOI: 10.48550/ARXIV.1701.05779

   6
9. Park, S.; Kim, H.-T.; Lee, S.; Joo, H.; Kim, H. Survey on Anti-Drone Systems: Components, Designs, and Challenges. *IEEE Access* **2021**. DOI: 10.1109/ACCESS.2021.3065926
10. Kümmritz, S.; Paul, L. Comprehensive Database of Drone Sounds for Machine Learning. Proceedings of the Forum Acusticum 2023, Turino, Italy, 11-09-2023, 667–674. DOI: https://dael.euracoustics.org/confs/fa2023/data/articles/000049.pdf10.61782/fa.2023.0049
11. Easy Access Rules for Unmanned Aircraft Systems (Regulations (EU) 2019/947 and 2019/945). Available online: https://www.easa.europa.eu/en/document-library/easy-access-rules/easy-access-rules-unmanned-aircraft-systems-regulations-eu (accessed on 05-01-2024).
12. Marcus, G. Deep Learning: A Critical Appraisal. *arXiv* **2018**. DOI: 10.48550/arXiv.1801.00631
13. Nanni, L.; Maguolo, G.; Paci, M. Data augmentation approaches for improving animal audio classification. *Ecological Informatics* **2020**, *57*, 101084. DOI: 10.1016/j.ecoinf.2020.101084
14. Oikarinen, T.; Srinivasan, K.; Meisner, O.; Hyman, J.B.; Parmar, S.; Fanucci-Kiss, A.; Desimone, R.; Landman, R.; Feng, G. Deep convolutional network for animal sound classification and source attribution using dual audio recordings. *J. Acoust. Soc. Am.* **2019**, *145*, 654–662. DOI: 10.1121/1.5087827
15. GitHub Repository, H2 Think gGmbH, DroneClassifier. Available online: https://github.com/H2ThinkResearchInstitute/DroneClassifier (accessed on 05-01-2024).

16. GitHub Repository, tensorflow, models, vggish. Available online: https://github.com/tensorflow/models/tree/master/research/audioset/vggish (accessed on 03-01-2024).
17. Shi, L.; Ahmad, I.; He, Y.-J.; Chang, K.-H. Hidden Markov model based drone sound recognition using MFCC technique in practical noisy environments. *J. Commun. Netw.* **2018**, *20*, 509–518. DOI: 10.1109/JCN.2018.000075
18. Xylo: Ultra-low power neuromorphic chip | SynSense. Available online: https://www.synsense.ai/products/xylo/ (accessed on 05-01-2023)
19. Branding, J.; Von Hörsten, D.; Wegener, J.K.; Böckmann, E.; Hartung, E. Towards noise robust acoustic insect detection: from the lab to the greenhouse. *KI - Künstliche Intelligenz* **2023**. DOI: 10.1007/s13218-023-00812-x
20. Sharma, G.; Umapathy, K.; Krishnan, S. Trends in audio signal feature extraction methods. *Applied Acoustics* **2020**. DOI: 10.1016/j.apacoust.2019.107020